HOMOSCEDASTICITY AND HETEROSCEDASTICITY

→ Homoscedasticity also known as homogenity of Variance/Assumption of equal variance. Homoscedasticity means same variance.

→ Homoscedasticity means having the same scatter. For it to exist in a set of data, the point must be about the same distance from the line.

→ Opposite is heteroscedasticity ("different scatter"), where points are at widely varying distance from regression line.

Homoscedasticity   Heteroscedasticity
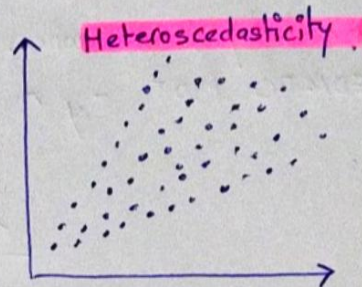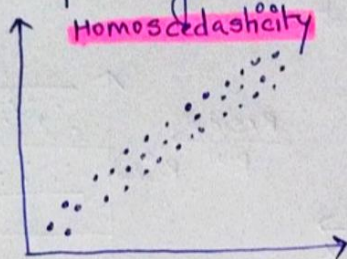
● Example of homoscedasticity →
- For example, suppose we want to explain student test score using amount of time each student spent studying.

   Dependent Variable → Test Score   Independent Variable → Time spent studying
- Error term would show the amount of variance in the test score that was not explained by amount of studying.

   Homoscedasticity → If the variance is uniform, it is homoskedastic and it would suggest that model may be adequate for test performance-explaining in terms of time spent studying.

   Heteroskedasticity → If the variance is not uniform, it is heteroskedastic.
   A plot of the error term may show a large amount of study time corresponds very closely with high test score but that low study time test scores varied widely and even included in some very high score. So variance of score may not be explained fully by study time.

● Further investigation may reveal that some student had seen the answer of question just before exam or they had taken similar test.
So to improve regression model, researcher would therefore add other variables indicating whether the student seen ans prior to test, same questions in previous exams etc.

- Heteroscedasticity is present when the size of the error terms differs across values of an independent variable.

→ Another example of heteroscedasticity →
   ° If we model household consumption based on income, we will find variability in consumption increase as income increase. When family have low income, they don't spend much on luxury items but in case of high income family, some family spend large amount on luxury item while some are moderate in luxury spending. So the variance is not uniform across families hence heteroskedastic.

DIAGRAM - Examine a scatterplot of the residuals against predicted value of dependent variable across a period of time.
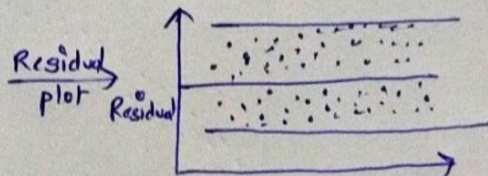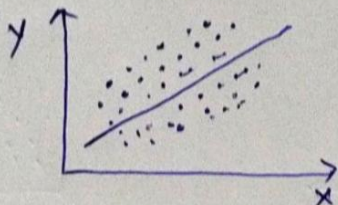
EXAMPLE — Sydney property sale price vs plot size
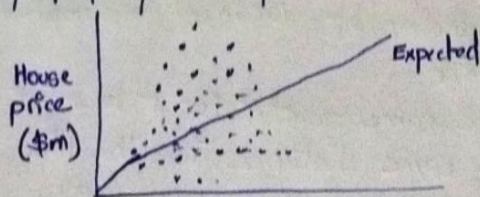
Price may vary due to location, essential shop near it.

Heteroskedasticity occurs when $V(Y|X)$ is not constant.

HOMOSKEDASTIC PLOT

In Regression, error term (Residual plot) should be homoskedastic.
In regression output, heteroskedasticity don't affect coefficient but heteroskedastic affect Standard error, p values and t values.

DETECTION OF HOMOSKEDASTIC

Detection 1 — Residual plot

Detection 2 — Goldfeld — Quandt test

Step 1 — Order the data by Independent variable (plot size)

Step 2 — Split the data. into two segment

In this we have A & B.

Step 3 — Run separate regression on each segment i.e, A & B.

The two regression line may not meet due to different data point.
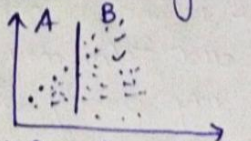
step 4 — Compare Variance estimate
If Variance $_A$ = Variance $_B$
then homoskedasticity

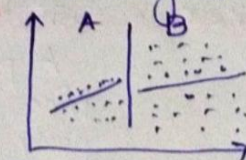Variance $_A \neq$ Variance $_B$ then heteroskedasticity.

$$F = \frac{MSE_A}{MSE_B}, \quad MSE \rightarrow Mean\ Square\ Error.$$
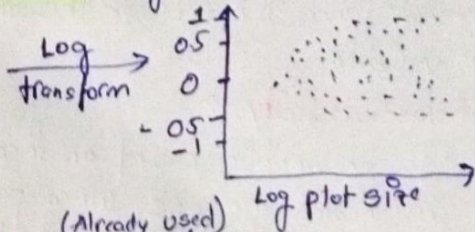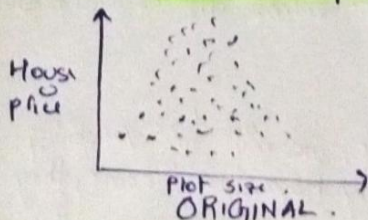
F value we find from this.

REMEDY FOR HETEROSKEDASTICITY

Best is LOG transformation

ORIGINAL. (Already used)

First few may be exception but the plot become more homoskedastic.

Real time example → ① Expenditure as a function of income ② income as function of age, ③ Speed measurement as a function of speed.